

P-19

NASA Case No.: MSC-22093-1

Print Figure: 1

PAT. APPL.  
IN-82  
157381

NOTICE

The invention disclosed in this document resulted from research in aeronautical and space activities performed under programs of the National Aeronautics and Space Administration. This invention is owned by NASA and is, therefore, available for licensing in accordance with the NASA Patent Licensing Regulations (14 CFR 12542.2).

In encouraging commercial utilization of NASA-owned inventions, it is NASA policy to grant licenses to commercial concerns. Although NASA encourages nonexclusive licensing to promote competition and achieve the widest possible utilization, NASA will consider the granting of a limited exclusive license, pursuant to the NASA Patent Licensing Regulations, when such a license will provide the necessary incentive to the licensee to achieve early practical application of the invention.

Address inquiries and all applications for license of this invention to NASA Patent Counsel, Lyndon B. Johnson Space Center, Code AL3, Houston, TX 77058. Approved NASA forms for application of nonexclusive or exclusive license are available from the above address.

Serial Number: 08/014,985

Date Filed: February 5, 1993

(NASA-Case-MSC-22093-1) METHOD AND  
APPARATUS FOR FILTERING VISUAL  
DOCUMENTS Patent Application  
(NASA) 19 p

N93-22017

Unclass

G3/82 0157381

## METHOD AND APPARATUS FOR FILTERING VISUAL DOCUMENTS

Visual documents may be defined as visual sequences on a recording media such as film, videotape or digital recordings. Visual documents are used to record event, for example, a motion sequence of the space shuttle in orbit displaying or retrieving a satellite. NASA stores a large amount of visual documents relating to operation of the space shuttle as well as various other manned and unmanned launches. Typically, such visual documents comprise many hours of film or videotape comprising a large number of frames and, frequently, very little change in the captured image occurs. For example, a visual document of the space shuttle in orbit may include a number of minutes of even hours of the robot arm positioned a certain way with very little movement. Methods which reduce the number of frames comprising a visual document to a lesser number of "key frames" are referred to as visual filtering. This procedure may also be referred to as abstracting the visual document. A "key frame" may be described as a frame in a visual document where a certain amount of motion or action, i.e., change, has occurred since the previously saved key frame. Although the application of visual documentation techniques has expanded in the last decade, methods for summarizing these documents have remained bound by human editing procedures. Such procedures are typically subject to high costs as well as variations and biases introduced by individual editors possessing different training backgrounds and aesthetic temperaments.

The present invention comprises a method and apparatus for producing an abstract of a video sequence of images, i.e., a visual document. The method assumes that video images viewed frame by frame change very little, and thus the video sequence is first sampled to produce a collection of frames which captures the essence of the footage. The sampled frames are then digitized and subjected to a structural decomposition process that reduces all information in each frame to sets of values. In the structural decomposition, selected features or parameters of each frame are decomposed into histograms, and the histogram of each feature is then converted to a single value by performing a Lorenz transform. These values are in turn normalized and then summed to produce only one information content value per frame. The information content for each frame is compared to a selected normal distribution cutoff point, and those frames having values greater than the cutoff point are selected. By selecting only those values at the tails of the normal distribution, key frames are filtered from their surrounding frames. Each of the selected frames are then compared with the value from the previous frame, and, if the values are not significantly different, the latter frame is not kept. The selected frames are then stored on a respective media.

Novelty is believed to exist in a method and an automated means for reducing a collection of video frames to a representative set of still frames which may be used in a video catalog or act as a video abstract to summarize the event recorded. The method according to the present invention can filter or compress a visual document with a reduction in digital storage on the ratio of up to 700 to 1 or more, depending on the visual document being filtered.

Inventor: Mark E. Rorvig and Robert O. Shelton

Employer: Johnson Space Center

Evaluator: Robert Savely

MSC-22093-1  
Serial No.: 08/014,985  
Filed: 2/5/93

## METHOD AND APPARATUS FOR FILTERING VISUAL DOCUMENTS

### ORIGIN OF THE INVENTION

The invention described herein was made by employees of the United States Government and may be manufactured and used  
5 by or for the Government of the United States of America for governmental purposes without the payment of any royalties thereon or therefor.

### FIELD OF THE INVENTION

The present invention relates to a form of data  
10 compression referred to as visual filtering, and more particularly to a method for extracting key frames from a visual document where change in the displayed image occurs to significantly reduce the number of frames comprising the visual document.

### 15 DESCRIPTION OF THE RELATED ART

Visual documents may be defined as motion or action sequences, i.e., visual sequences, on a recording media such as film, videotape, or digital recordings. Visual documents are used to record a certain event or events. For example,  
20 the visual document may contain a motion sequence of the space shuttle in orbit obtaining or retrieving a satellite, etc. Visual documents constitute a major source of information for various government agencies and private sector entities. For example, the National Aeronautics and Space Administration  
25 (NASA) stores a large amount of visual documents relating to operation of the space shuttle as well as various other manned and unmanned launches. Typically, such visual documents comprise many hours of film or videotape comprising a large number of frames where very little change in the captured

shuttle in orbit obtaining or retrieving a satellite may include a number of minutes or even hours of the space shuttle's robot arm positioned a certain way with very little movement. Methods which reduce the number of frames  
5 comprising a visual document to a lesser number of "key frames" are referred to as visual filtering. This procedure may also be referred to as abstracting the visual document. A "key frame" may be described as a frame in a visual document where a certain amount of motion or action, i.e., change, has  
10 occurred since the previously saved key frame, such that this key frame would also be stored in the filtered visual document.

Although the application of visual documentation techniques has expanded in the last decade, methods for  
15 summarizing these documents have remained bound by human editing procedures. Such procedures are typically subject to high costs as well as variations and biases introduced by individual editors possessing different training backgrounds and aesthetic temperaments. While there currently exist many  
20 different image decomposition techniques, there has been no attempt to abstract or index visual documents using visual parameters directly. Therefore, a method and apparatus is desired to automatically filter visual documents using various selected visual parameters. There has been a long felt need  
25 for such a visual filtering method and apparatus.

U.S. Patent No. 6,060,290 to Kelly et al. discloses a gray scale automated visual analysis system which is used in the grading of fruits and nuts. A video inspection system obtains a video image of the produce as it passes by on a  
30 conveyor belt. The video image is decomposed into a representative histogram which identifies characteristic features of interest, such as color, shape, size, bruising, etc. The histogram is normalized and then various attributes

are derived from the histogram. These attributes are then assigned various codes to aid in determining whether the produce is acceptable or unacceptable. A sorting apparatus diverts the unacceptable produce as a result of the selection  
5 process.

U.S. Patent No. 5,103,307 to Sugiyama discloses an interframe coding scheme which composes one image frame from a previous frame and performs a differencing operation between the two. The differences are stored to reduce the storage  
10 necessary for a collection of frames which are similar. Significant differences are identified as larger than average due to the fact that many differences in the images exist (scene change). U.S. Patent No. 4,937,685 to Barker et al. teaches a video composition apparatus in which image source  
15 material from a plurality of sources are selected and then connected to form a program sequence. Intended applications include archive news footage and the like that require rapid identification and playback. A computer is used to control independent video sources to provide a quick scan capability  
20 for search purposes. The user interface allows the marking of start frames and end frames and also permits the concatenation of frame sequences to an output device.

U.S. Patent No. 5,012,334 to Etra relates to a video editing system for organizing a collection of video archives  
25 stored on video disks. A computerized editing system controls access to footage stored on the video disks by means of a keyword searchable index stored on the computer's hard drive. A query access language allows the user to prepare a script which identifies related segments in the order prescribed by  
30 the script. Additional features support other editing features as well as special effects.

SUMMARY OF THE INVENTION

The present invention comprises a method and apparatus for producing an abstract of a video sequence of images, i.e., a visual document. The method assumes that video images  
5 viewed frame by frame change very little, and thus the video sequence is first sampled to produce a collection of frames which captures the essence of the footage. The sampled frames are then digitized and subjected to a structural decomposition process that reduces all information in each frame to sets of  
10 values. In the structural decomposition, selected features or parameters of each frame are decomposed into histograms, and the histogram of each feature is then converted to a single value by performing a Lorenz transform. These values are in turn normalized and then summed to produce only one  
15 information content value per frame.

The information content value for each frame is compared to a selected normal distribution cutoff point, and those frames having values greater than the cutoff point are selected. By selecting only those values at the tails of the  
20 normal distribution, key frames are filtered from their surrounding frames. Each of the selected frames are then compared with the value from the previous frame, and, if the values are not significantly different, the latter frame is not kept. The selected frames are then stored on a respective  
25 media.

Therefore, the present invention provides an automated means for reducing a collection of video frames to a representative set of still frames which may be used in a video catalog or act as a video abstract to summarize the  
30 event recorded. The method according to the present invention can filter or compress a visual document with a reduction in digital storage on the ratio of up to 700 to 1 or more, depending on the visual document being filtered.

**BRIEF DESCRIPTION OF THE DRAWINGS**

A better understanding of the present invention can be obtained when the following detailed description of the preferred embodiment is considered in conjunction with the following drawings in which:

Figure 1 illustrates a system for filtering visual documents according to the present invention;

Figures 2A, 2B and 2C are flowchart diagrams illustrating operation of the visual document filtering method according to the present invention; and

Figure 3 illustrates the process of selecting key frames using a normal distribution cutoff point.

**DETAILED DESCRIPTION OF THE SPECIFIC EMBODIMENT**

Referring now to Figure 1, a system for filtering visual documents according to the present invention is shown. The system includes a first recording medium, for example, a video cassette recorder referred to as VCR1, which stores the visual document to be filtered. VCR1 includes an output connected to an input of a second VCR referred to as VCR2, which is used to store the filtered visual document. VCR1 includes another output connected to a frame buffer and digitizer 24 which in turn has an output connected to a computer 22. The computer 22 has outputs connected to VCR1 and VCR2 to control their operation. The computer 22 implements the method according to the present invention to filter or abstract visual documents to a much smaller number of frames. The computer 22 first reviews the visual document, performs the method according to the present invention to select key frames and then directs VCR1 to transmit these key frames to VCR2. Various other arrangements of components can be utilized, depending on the medium of the visual document, the medium of the abstracted document and the storage capabilities of the computer system.

Referring now to Figures 2A, 2B, and 2C, a flowchart diagram illustrating operation of the method according to the present invention is shown. The method according to the present invention can be implemented on various types of hardware, including a general purpose personal computer 22 as  
5 illustrated in Figure 1 or dedicated video processing logic, as necessary.

In step 102, the method performs an initial sampling of frames comprising the visual document. As an example of the  
10 operation of this step, consider a visual document one or more hours in length comprised of 16 or more frames per second of interleaved video or film. In each second, little change among any of the frames occurs. The method assumes that video images viewed frame by frame change very little, and thus the  
15 visual document is sampled initially to produce a smaller collection of frames which captures the essence of the visual document. In one embodiment, a sampling rate of one frame per every one second constitutes an adequate collection of sampled frames for filtering. In another embodiment, the sampling  
20 rate of one frame of video imagery per every five seconds constitutes an adequate collection of sampled frames which may then be further filtered. It is noted that the sampling frequency will necessarily depend on the type and amount of motion sequence in the respective visual document being  
25 filtered.

The sampling in step 102 can occur prior to the visual document being stored in VCR1, or can be performed by the computer 22 while the document is present in VCR1. By initially sampling the document in this manner, the number of  
30 frames that are required to be processed in the remaining steps is greatly reduced. Having sampled the visual document to extract a number of frames, for example, sampling at one frame per second to produce N total frames, the method then



digitizes each sampled frame in step 102 if the sampled frames are in analog form. The frames are digitized by the frame buffer and digitizer 24, preferably at the direction of the computer 22.

5        In step 104, the method decomposes certain selected features or parameters of each sampled frame into a number of histograms. In other words, each of the sampled frames is processed to extract a number M of general primitive features rendered as histograms. For example, in one embodiment, the  
10 parameters of pixel intensity, edge intensity, edge slope, line length, line distance from image origin, and angles for each frame are extracted from each frame using histograms. The pixel intensity histogram accumulates gray scale values in 64 intervals. The histogram values for edge intensity, edge  
15 slope, line length, line distance from image origin and angles are calculated after performing a Hough transform on each respective frame.

      In the preferred embodiment, edge intensity is defined as a constant gray scale value of greater than 5 pixels in width  
20 and accumulated in 64 histogram intervals. Edge slopes are preferably accumulated in 45 histogram intervals of 2 degrees each. Lines are defined as constant gray scale values of less than 5 pixels and accumulated in 64 histogram intervals of 4 pixels to 256 pixels. Line distance from the origin is  
25 preferably defined as the number of pixels between the center of the respective line and the largest value of XY coordinates of the image. Line distance from the origin is preferably accumulated in 64 histogram interval of 4 pixels to 256 pixels. Angles of lines are accumulated in 45 histogram  
30 intervals of two degrees each.

      The algorithms for obtaining these histograms are well known in the art. For more information on these algorithms, please see Ballard, D.H. and Brown, C.M. "Computer Vision,"

Prentice-Hall, Englewood Cliffs, NJ, 1982. See also Leigh, Albert "A Fuzzy Measure Approach to Motion Frame Analysis for Video Data Abstraction" 1992, a Masters Thesis available at the University of Houston-Clear Lake. It is noted that other  
5 features, instead of or in addition to the above parameters, may be used. For example, in an alternate and preferred embodiment, the parameters of hue, chroma, and saturation of each pixel are substituted for gray scale pixel intensity.

As another example of additional features, it should also  
10 be noted that wavelet based coefficients could be used to provide histograms of image features which would produce very similar measures as those produced in the preferred embodiment. These measures are described in Mallat, Stephan, "A Theory of Multiresolution Signal Decomposition - the  
15 Wavelet Decomposition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 11, pps. 674-693, 1989.

Further, still another method of obtaining feature histograms would be the use of histogramal power spectrums from fourier transforms, commonly known as FFT's. An  
20 advantage of FFT implementation for obtaining the histograms would be that specialized computer hardware already exists for producing these measures. The article by Young, Tzay Y. and Liu, Philip S., "VLSI Array Architecture for Pattern Analysis and Image Processing," in Handbook of Pattern Recognition and  
25 Image Processing, Academic Press, pps. 471-496, 1986 describes the use of hardware to obtain FFT and other histogram measures by custom fabrications of hardware.

In step 106, the method converts each of these feature histograms into a single value by performing a Lorenz  
30 transform on each respective histogram. By converting the respective histograms for each image into Lorenz values, the information for each frame is reduced to a small set of real numbers where each number constitutes a structural attribute

of the entire frame. Therefore, for N total frames being examined and for M parameters, the result is an N x M array of values. For more information on the use of the Lorenz transform, please see Chang, S.K. and Yang C.C. "Picture  
5 Information Measures for Similarity Retrieval," Computer Vision, Graphics, and Image Processing, 23:366-375, 1983.

As illustrated in Figure 3, the filtering method of the present invention selects frames in part based on the relative position of each frame in the normal distribution of all  
10 sampled frames in the collection being examined. Therefore, each individual structural feature of the frames is assumed to be normally distributed. This assumption may be incorrect however, depending upon the composition of any set of particular frames decomposed by any particular feature. To  
15 overcome the weaknesses of this assumption, the structural features of each frame are themselves averaged by calculating the mean and standard deviation for each Lorenz value across the sampled frames in step 108 and converting the individual Lorenz values in each frame to unit normal deviations of the  
20 normal curve in step 110. Therefore, the method in step 108 calculates the standard deviation and mean for each of the M decomposition values or Lorenz values across each of the N sampled frames. In step 110, the method converts each of the M Lorenz values of each of the N frames into a unit normal  
25 deviate from the normal distribution of that feature using the respective means and standard deviations calculated in step 108. Thus all values are reduced to a common unit of measurement.

In step 112, the method calculates the sum of the M unit  
30 normal deviate values for each frame, reducing the total information content of each respective frame to a single value. This produces N values, i.e., one value for each respective frame. These values are referred to as information

content values. By simply summing all of the values, each frame is represented by a single value encompassing its entire structure. In step 113, the method calculates the mean and standard deviation of the information content values across  
5 all of the selected frames.

In step 114 (Figure 2B), the method selects a distribution cutoff point from the normal distribution, for example, 1.2 in one embodiment. Preferably, the distribution cutoff point is determined using the mean and standard  
10 deviation calculated for all of the information content values in step 113. For example, in the preferred embodiment the distribution cutoff value is selected based on a given or selectable number of standard deviations from the mean. Alternatively, the distribution cutoff point is user  
15 programmable.

In step 116, the method first normalizes the information content value of the frame being examined using the mean and standard deviation calculated in step 113. This is similar to the normalization that occurred in step 110. The method then  
20 compares the normalized information content values of the frame to the set distribution cutoff point and determines whether the normalized information content value for the respective frame being examined is greater than the cutoff point in step 118. If so, the frame is stored in step 122.  
25 If not, the frame is discarded in step 120. Upon completion of either steps 120 or 122, the method determines if the last frame has been examined in step 124. If not, the method returns to step 116 to examine the next frame. If the last frame has been examined in step 124, then the method advances  
30 to step 130. Therefore, steps 116-124 operate to select those frames having a normalized information content value greater than the cutoff point chosen in step 114. The above steps are

essentially equivalent to selecting the frames at the tails of a normal distribution as illustrated in Figure 3.

In step 130, the method, beginning with the second selected frame, compares the respective frame's value with the value from the previous selected frame. If the frame value is determined to not be significantly different from the prior frame's value in step 132 (Figure 2C), then the frame is discarded in step 134. In step 136, which follows step 134 or step 132 if the frame is to be kept, the method determines if the current frame being examined is the last frame. If not, the method returns to step 130 (Figure 2B) to continue examining the remaining frames. If the last frame has been examined in step 136, then the method advances to step 138. Therefore, steps 130-136 serve to further reduce the number of frames comprising the filtered document. It is noted that, in an alternate embodiment, steps 130 - 134 can follow step 122 to reduce required storage space.

In step 138, the selected frames are output to appropriate digital or analog recording media, VCR2 in the illustrated embodiment, which stores the filtered video document.

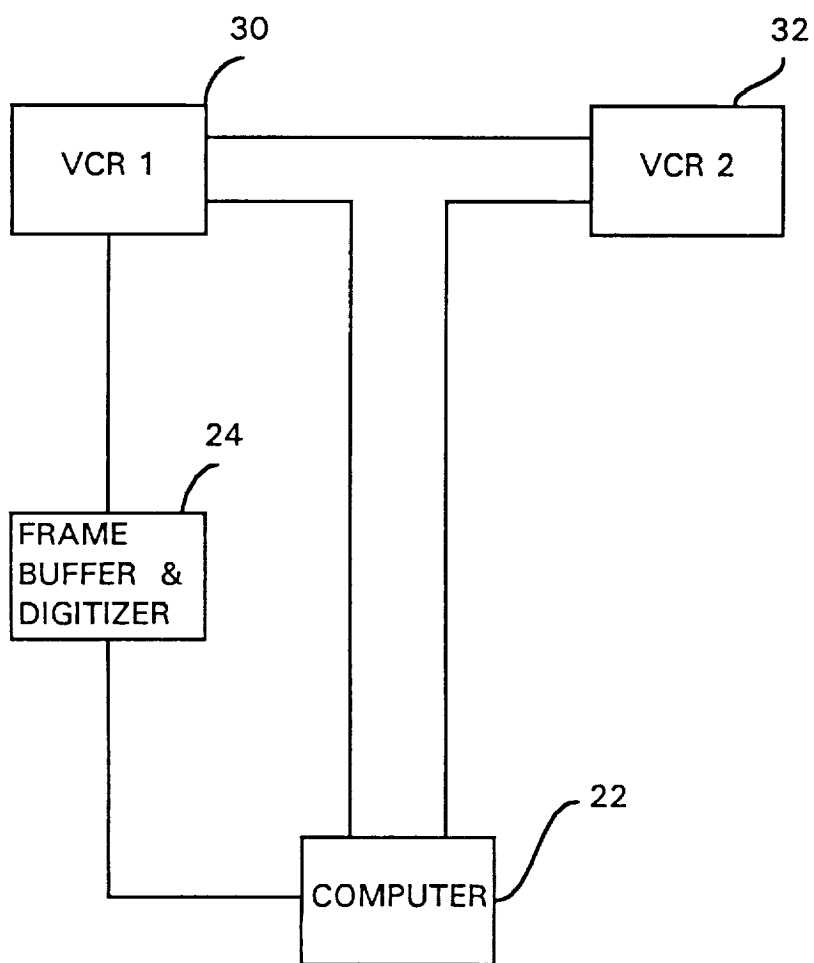
Therefore, a method is provided for automatically filtering key frames from visual documents to provide a much smaller visual document that represents or abstracts the motion sequence of the original visual document. The frames comprising the visual document are first sampled at a selected rate and then digitized. These sampled and digitized frames are then subjected to a structural decomposition process that reduces all information to a set of values for each frame. These values are in turn normalized and further combined to produce one information content value per frame. These information content values are fitted to a normal distribution of all values in the respective set of frames. By selecting

only those values at specified areas at the tails of the distribution, i.e., above a selected distribution cutoff point, the key frame images are filtered from their surrounding frames.

- 5       The foregoing disclosure and description of the invention are illustrative and explanatory thereof, and various changes in the components and steps, as well as in the method of operation may be made without departing from the spirit of the invention.

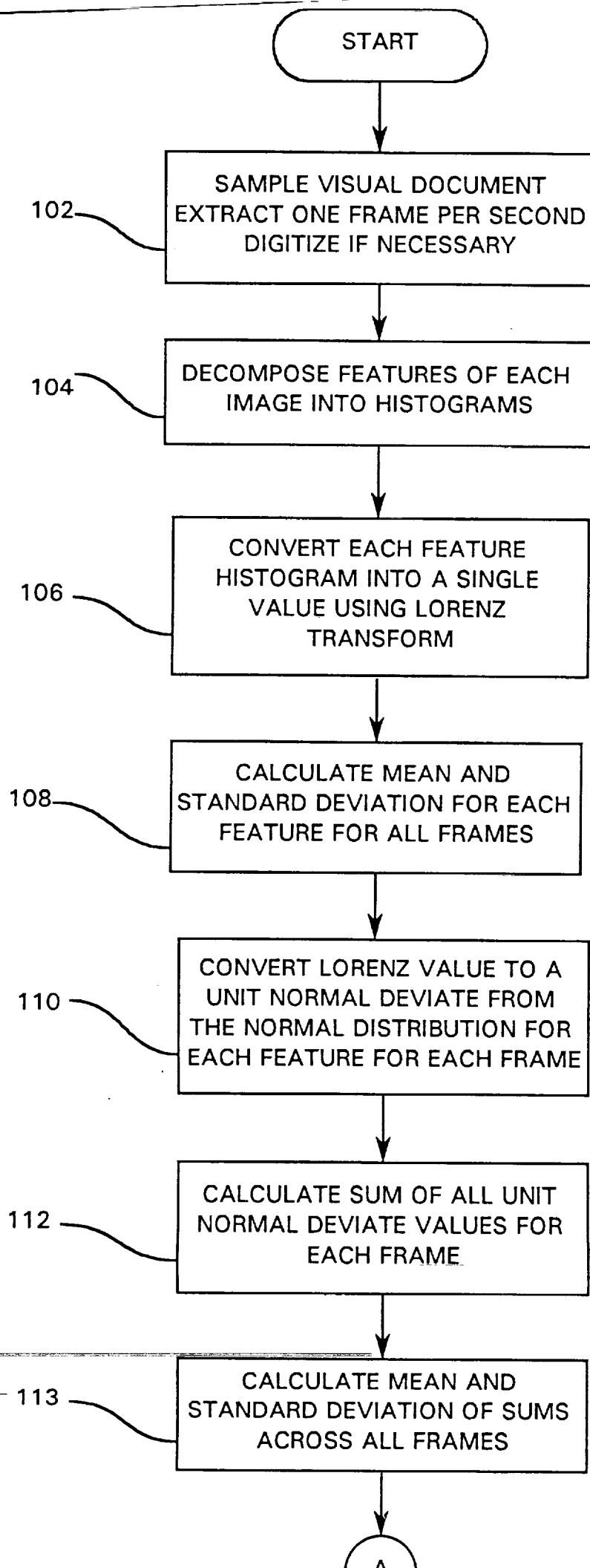
ABSTRACT OF THE INVENTION

A method and apparatus for producing an abstract or condensed version of a visual document. The frames comprising the visual document are first sampled to reduce the number of frames required for processing. The frames are then subjected to a structural decomposition process that reduces all information in each frame to a set of values. These values are in turn normalized and further combined to produce only one information content value per frame. The information content values of these frames are then compared to a selected distribution cutoff point. This effectively selects those values at the tails of a normal distribution, thus filtering key frames from their surrounding frames. The value for each frame is then compared with the value from the previous frame, and the respective frame is finally stored only if the values are significantly different. The method filters or compresses a visual document with a reduction in digital storage on the ratio of up to 700 to 1 or more, depending on the content of the visual document being filtered.

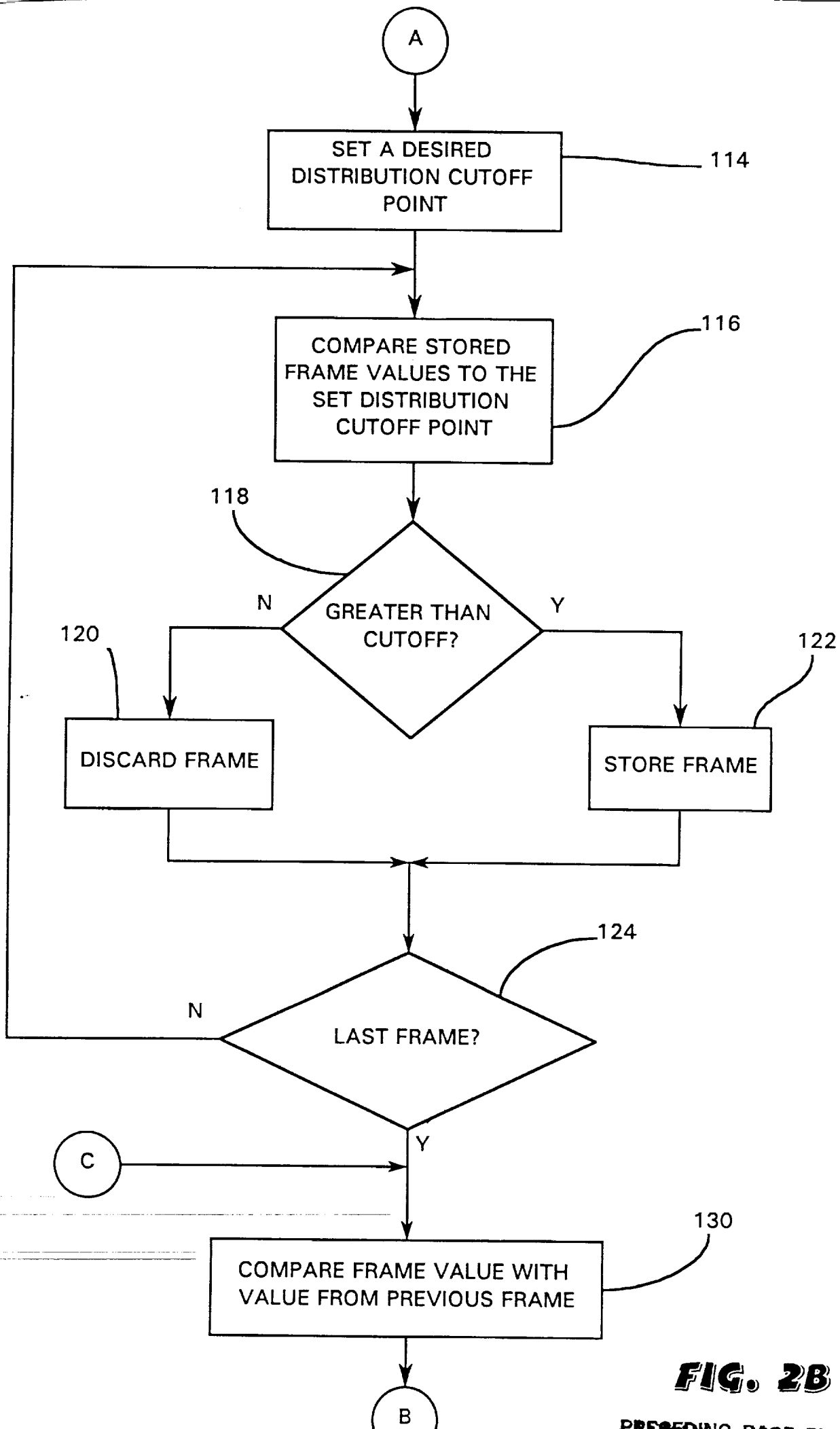


**FIG. 1**



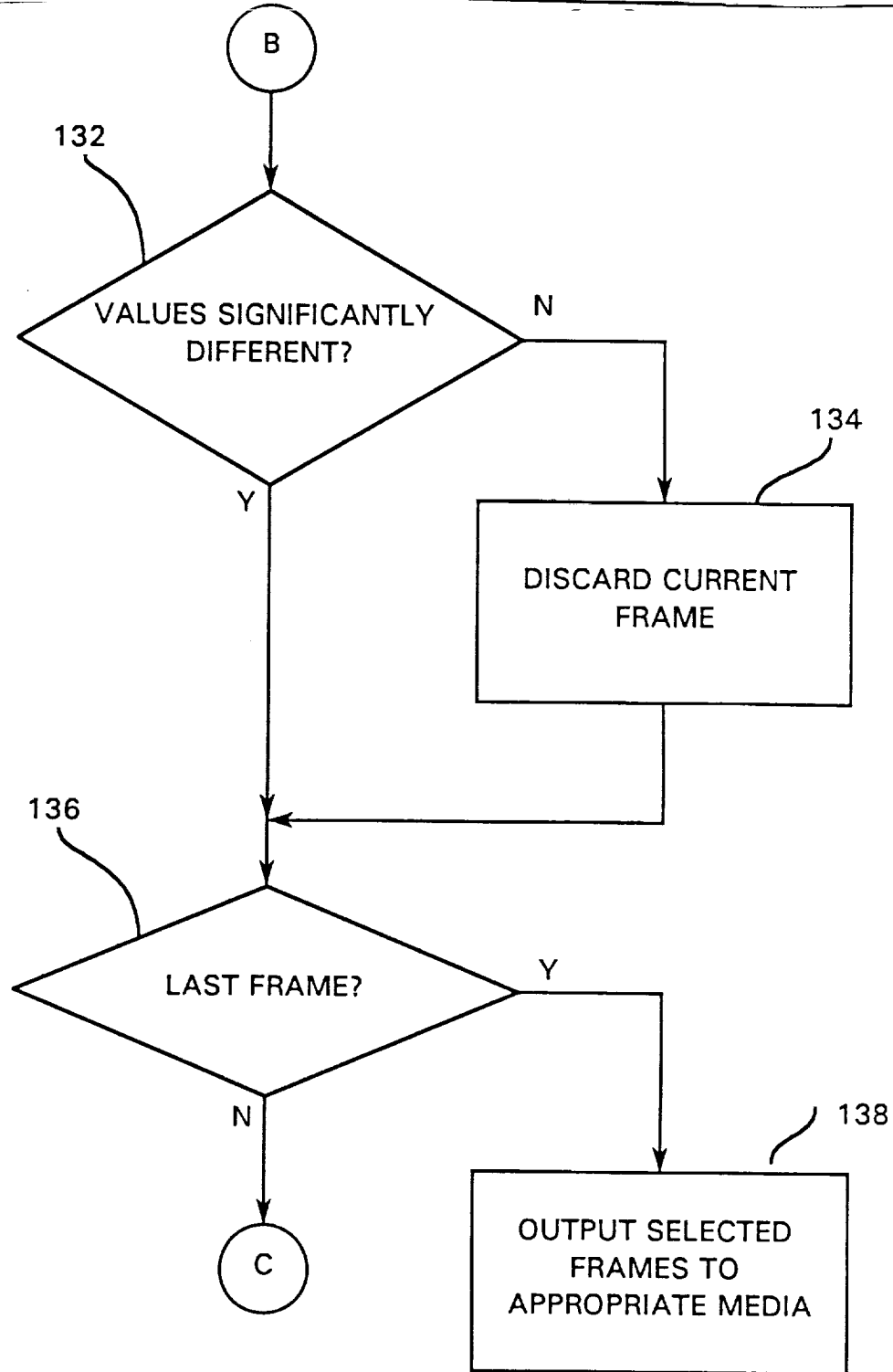


**FIG. 2A**

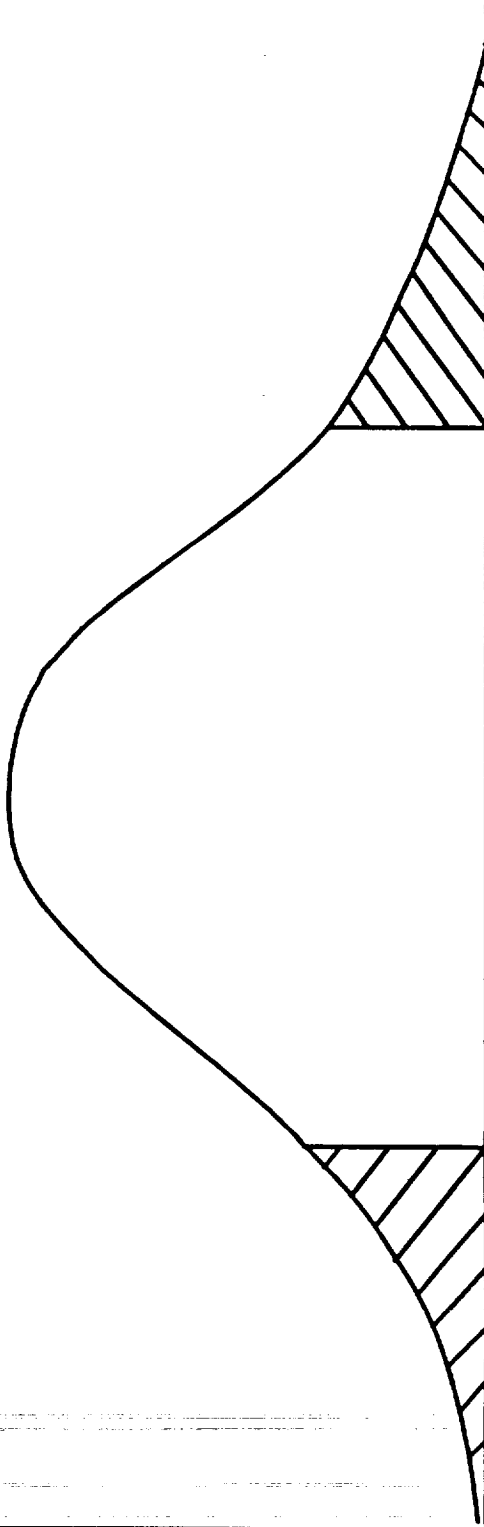


**FIG. 2B**

**FIG. 2C**



PRECEDING PAGE BLANK NOT FILMED



**FIG. 3**